

Figure 1

The diagram illustrates a speech processing system divided into two main sections: Non-real-time and Real-time, separated by a vertical dashed line.

Non-real-time section:

- sound** (60) is input to **spectral estimation** (62).
- spectral estimation** (62) leads to **pitch detection** (64).
- pitch detection** (64) produces **pitch estimates**.

Real-time section:

- sound** (60) is also input to **spectral estimation (new window length)** (66).
- spectral estimation (new window length)** (66) leads to **peak detection** (68).
- pitch estimates** from the Non-real-time section are fed into **peak detection** (68) via a dashed arrow.
- peak detection** (68) produces **peak estimates**.
- peak estimates** lead to **smoothing** (70).
- smoothing** (70) leads to **overlap-add frames**.
- overlap-add frames** lead to **manipulation/control** (72).
- manipulation/control** (72) leads to **additive synthesis** (74).
- additive synthesis** (74) produces the final **sound** (76).

Fig. 2

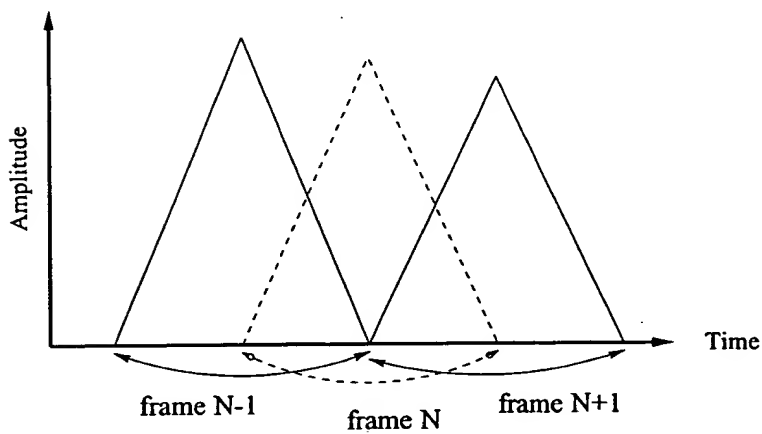


Fig. 3

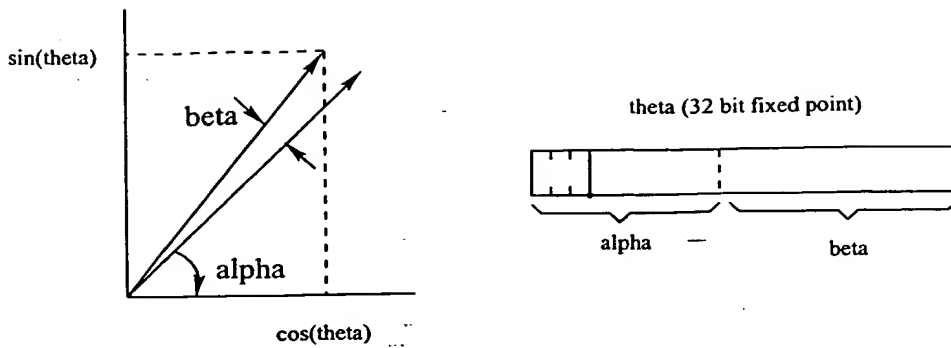


Fig. 4

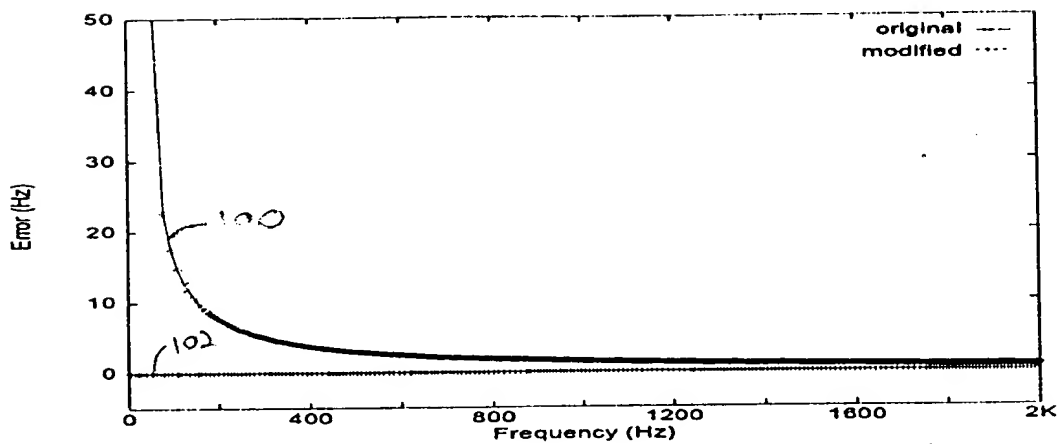


Fig. 5

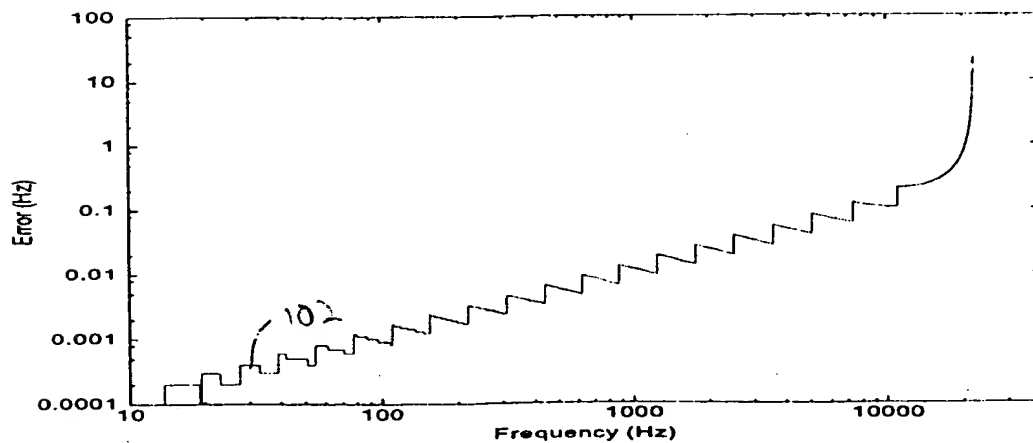


Fig. 6